This application is submitted in the names of John D. Marshall and John C. Gaddy and has been assigned to Timbral Research Inc.

SPECIFICATION

COMPUTER BASED AUTOMATIC AUDIO MIXER

FIELD OF THE INVENTION

The present invention relates to an apparatus and a method for mixing at least two audio files. More specifically, the apparatus and methods of the present invention enable a user to achieve a professional quality sound recording without having any recording engineering training or experience.

1. PRIOR ART

Mixing of recorded audio programs has been performed since the advent of multiple audio track recording. Multiple track recording allows a user to record an

audio performance onto a single piece of media, though each of the tracks is completely independent from one another. For example, in a two track recording the vocal track may be separately recorded onto one track while the remaining performance would be recorded onto the other track.

In order to create a multiple track recording special equipment and the knowledge of how to use the equipment is required. Typically, a recording engineer is employed to run the equipment and make the recording. An experienced recording engineer will be able to best utilize multiple track recording technology to create the best audio recordings possible.

For example, a recording engineer making a multiple track recording may record each of the tracks independently. The vocalist would be placed in the recording booth and an accompaniment track would be played back through a set of headphones so the vocalist could sing along with the track. The vocalist performs with the accompanying musical track, and the synchronization occurs naturally because the two tracks coexist on the same recording medium. After successfully making a multiple track recording, the recording engineer may apply

electronic processing to each individual track to adjust the overall characteristics of the entire multiple track recording, or master recording. This processing may include balancing the instruments, adding reverberation, equalization, audio compression, noise reduction and stereo imaging. After the processing is completed, the individual tracks are combined into a mixed down stereo or monaural master. In the stereo master, several instruments or voices are combined into a pair of channels to create a stereo image.

Traditionally, the mixing process has been accomplished by an analog electronic circuit, or mixer, comprising an array of amplifiers each with its own manually adjustable volume control. The circuit includes a single summing amplifier for monaural, or a pair of summing amplifiers for stereo to linearly combine the outputs of the channel amplifiers. The individual channel volume controls can be adjusted manually during the mixing process to adjust the levels of the instruments in the mix. Using this method, individual channels may be added or removed from the overall mix. Finally, additional effects may be applied to the final mix.

With the advancements in electronics, analog mixing boards have been automated. That is the sliders that are used to control the levels of each channel amplifier have been motorized and may adjust automatically. The sliders can be controlled with a memory and a playback unit that synchronizes the mixing board with the analog recording. This allows the final mixing scheme, including all variations of the slider positions over the duration of the recording to be arranged and recorded prior to making a master recording. The final mixing scheme may then be played back while recording the final mix.

The advancements described above have been applied to digital recording systems. Digital mixing boards function in the same manner as the analog boards described above. Though, instead of utilizing analog audio signals, digital mixers are capable of utilizing digitally recorded audio material. For example, traditional analog signals are digitized to create audio files that are stored onto a computer hard drive or onto a magnetic tape or another digital storage medium. Individual mixing levels may be adjusted manually, or the mixing board may be automated as described above to reflect the manual adjustments made to the mix.

Each of the systems described above requires expensive hardware that is difficult to operate and is expensive to maintain. In order to fully utilize the functions of a mixing board, a recording engineer must have a great knowledge of the functions of the mixing board and the affect that each change will have on the overall sound of the master recording. Also, existing automated mixing systems require mixing levels to be set by the recording engineer before they can be automatically played back.

Additionally, an artist will often rent studio time in order to make a recording. Artists may themselves be capable recording engineers, but in order to make a recording the artist would have to function as both the recording engineer and the performing artist, which is very difficult, if not impossible. Therefore, in addition to renting the studio, an artist will typically employ a recording engineer to run the mixing board during the recording process, which increases the cost of making a recording.

A recent variation on the mixing methods described above has been the advent of software mixing and audio recording programs that can be run on a

personal computer. As the processing power of personal computers has advanced so has the ability to utilize a computer for the mixing necessary to make a master recording. For example, a personal computer running Microsoft Windows® operating system and any one of the following audio mixing programs such as Pro Tools from Digidesign, or Vegas and Sound Forge available from Sonic Foundry, or Cool Edit Pro available from Syntrillium, or Cubase available from Steinberg can replace digital mixing boards in a recording studio. Though the personal computer software can be utilized to lower the costs of making a master recording by eliminating multiple dedicated hardware devices in a recording studio, the presently available mixing programs are still very expensive.

Also, the digital computer-based mixing programs mentioned above require an extraordinary amount of skill and knowledge to operate. Not only does the user have to be an experienced recording engineer, the user must also be able to configure a personal computer to use the mixing programs. Furthermore, many of the programs listed above include extensive user manuals, which must be read and understood before a user can maximize the performance of the software.

Moreover, understanding the manuals often requires training classes and advice from customer support engineers.

A recording and mixing system is a useful tool for learning to play a musical instrument and for learning a foreign language. If a music student has an opportunity to play along with musical accompaniment and can quickly hear back a professional quality mix of his or her performance with the accompaniment, the student can adjust her or his performance, try the piece again and progress is rapid. Similarly, foreign language students benefit when they can record a phrase and compare it to that of a native speaker. As described above, the audio mixing process is traditionally a difficult one and even if the student is a skilled recording engineer, attention to the technical details of the recording and mixing process diverts the student from the task of learning to play his or her musical instrument or learning to perform a foreign language dialogue.

Therefore there is a need for a recording and mixing system that simplifies the process described above to allow music students to produce high quality recordings while keeping their focus on the music.

There is also a need to facilitate an online language lab for foreign language students that offers a method and apparatus for performing a part in a foreign language dialogue and easily mixing it with the other part of the dialogue or mixing a phrase with a matching phrase from a native speaker.

Furthermore, the cost of the equipment necessary to provide such recording and mixing functions is far out of reach of a typical music student. Therefore, it is desirable that the proposed system could be implemented on a simple personal computer requiring only a minimal amount of training and cost to users.

A primary objective of this invention is to provide an automatic mixing system that emulates the listening, analysis and adjustment processes traditionally provided by the recording engineer. That is, the object of this invention is to provide an expert system to replace the recording engineer and associated hardware.

SUMMARY OF THE INVENTION

The present invention provides a method and apparatus that automatically mixes at least two digital audio files to produce a single output file as if it were produced by a recording engineer. The method and apparatus of the present invention allows a user to utilize a relatively inexpensive personal computer as a digital recording studio. This is accomplished by operatively coupling the personal computer with a more powerful server computer via an Internet (TCP/IP) or other digital communications connection. The server computer implements expert digital audio mixing functions comprising the following components, (1) a digital audio file reading and analysis program, (2) a digital audio summing program. Alternatively, the digital mixing program of the present invention may be installed on the client computer, though preferably the mixing program is disposed on the server computer as described above.

The present invention may be used to mix any number of digital audio files. However, for simplicity, the following discussion is limited to the mixing of two files. The first file is a pre-recorded accompaniment file residing on the server, and the second is a user-recorded digital audio file transmitted to the server by

software on the client computer system via a network connection. The user may have created the second digital audio file using the methods and apparatus in co-pending application entitled "SYNCHRONIZED STREAMED PLAYBACK AND RECORDING FOR PERSONAL COMPUTERS" having Serial No. XX/XXXX filed on December 27, 2000, and assigned to Timbral Research Inc, hereby incorporated in its entirety by reference. The co-pending application entitled "ONLINE COMMUNICATION SYSTEM AND METHOD FOR AURAL STUDIES" having Serial No. XX/XXXX, filed on December 27, 2000, and assigned to Timbral Research Inc, hereby incorporated in its entirety by reference. describes a learning system incorporating both the recording and mixing patents. Alternatively, the user may have created the second audio file utilizing any of the above mentioned programs. Furthermore, the user may have created the second audio file using other means as described in greater detail below.

If the user-recorded audio was made using an analog audio recorder, it would have to be digitized using one of several means known in the art. Alternatively, if the audio was captured using a digital audio recording device, such as a Digital Audio Tape (DAT) recorder, a hard drive recorder, or any other

digital audio recording device capable of creating a digital audio file, the digital audio file would then have to be transferred to and stored onto the client computer and transmitted to the server computer for use by the digital mixing program. The audio files may be in any format, as long as they may be read by the computer to produce simple time samples. The sample rates may differ and are converted as needed as part of the mixing process. If time alignment is critical then the starting points of each input file must possess the desired time correspondence so that after mixing they will be aligned correctly. The bit depth of the files may also differ; roundoff errors are avoided by implementing all of the computations using arithmetic with at least two (2) bits greater precision than the greatest bit depth among the input files. For example, if the highest precision file was digitized to 16 bits, then all the computations must be carried out with at least 18 bit precision.

After uploading the second digital audio file to the server, the digital mixing program reads and processes the two digital audio files twice. In the first pass the files are read and analyzed to determine scale factors to be used in the mixing process while the actual mixing is accomplished in the second pass.

The first pass is begun when the program reads the audio file headers to determine the file formats. If the digital audio files are in readable, non-compressed formats such as WAV or AU, no processing is performed at this step. However, if either or both of the files are in a compressed format such as MPEG-2 Layer III (MP3), Real Media (RM) or Quick Time (QT), the compressed file or files are expanded to a simple time sample format. At this point, all the samples from each file are processed by applying DSP routines to add audio compression, artificial reverberation, synthetic stereo imaging, etc. In this process, data are collected sample by sample for each file so that after all samples are processed, characteristic parameters are calculated for each file. Typically, these parameters include but are not limited to a peak absolute value and a root mean square (RMS) value for each processed audio file. In the case of a stereo input file or a stereo processed result from a monaural input file, the characteristic parameters are the result of examining the complete set of samples, including both the left and right channels. Alternatively, the DSP application may be bypassed during the first pass if its effect on the resulting peak absolute value and RMS value can be estimated accurately. A scale factor is then calculated for each digital audio file from their

respective peak absolute values and RMS values. The scale factors are stored for application in the second pass.

The second pass begins with a second reading of samples from the input audio files and the application of DSP functions, such as audio compression, artificial reverberation, or stereo imaging. Next, if the resulting audio data files possess differing sample rates, the lower rate file is converted up to the higher sample rate or the higher rate file is converted down to the lower sample rate. This is accomplished by one of many means commonly known in the art and may be done by simple linear interpolation if the sample rates differ by an integer multiple. The resulting samples from the two files are multiplied by their respective scale factors, and then time-corresponding samples that have been processed, converted and scaled are summed. Finally, the resulting single set of samples is written to produce a single digital audio output file. The output file contains a high quality audio result in which neither audio program dominates the mix and all samples have values within the acceptable range of the output file format. For example, if one input file has higher amplitude than the other, the file with the lower amplitude will be scaled up and the file with the higher amplitude

will be scaled down to normalize the amplitude of the overall mix. Still further, when mixing at least two audio files, if one file is greater in length than the other, during the mixing process the time length of the shorter audio file will be extended by appending zero-valued samples to the end of the file as necessary.

This invention further relates to machine readable media on which are stored embodiments of the present invention. It is contemplated that any media suitable for retrieving instructions is within the scope of the present invention. By way of example, such media may take the form of magnetic, optical, or semiconductor media. The invention also relates to data structures that contain embodiments of the present invention, and to the transmission of data structures containing embodiments of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGURE 1 is a high level function flow diagram illustrating the present invention.

FIGURE 2A is a functional flow diagram of the digital audio file reading and analysis program.

FIGURE 2B is a functional flow diagram of an alternative embodiment of the digital audio mixing program of the present invention.

FIGURE 2C is a functional flow diagram illustrating a second alternative embodiment of the digital mixing program of the present invention.

FIGURE 3A is an expanded diagram illustrating the calculation of the scale factors for two digital audio files.

FIGURE 3B is an expanded diagram illustrating the method for calculating scale factors for N audio files.

FIGURE 4 is an expanded functional flow diagram of the digital audio summing program.

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

Though the digital mixing program 90 of the present invention will be described below in reference to a monaural signal, this should not be considered limiting in any manner. Furthermore, digital mixing program 90, can be readily applied to stereo recordings. For example, in the following description, where reference is made to determining a peak value during the analysis process, the value would be determined for a stereo file from the entire set of input samples including both left and right channels.

Referring now to FIG. 1 there is shown a high level function flow diagram of the digital mixing program 90 of the present invention. As shown in FIG. 1, digital mixing program 90 is divided into two separate boxes, BOX 100 and BOX 200. Referring now to BOX 100, upon initiation of digital mixing program 90, at BOX 110 the header of the first audio file is read and it is determined whether the file is in a compressed format.

At Diamond 120, if the file is not in a compressed format, digital mixing program 90 continues to BOX 140. If the file is in a compressed format, digital mixing program 90 proceeds to BOX 130, the file is expanded, and the program 90 continues to BOX 140.

At BOX 140, samples from the file are read.

At BOX 150 the samples are pre-processed to add reverb, stereo imaging or other DSP effects.

At BOX 160 digital mixing program 90 determines the peak absolute value attained over the duration of the pre-processed audio file and the root mean square (RMS) average of the pre-processed sample values in the file.

At Diamond 170 digital mixing program 90 checks to see if there are any additional audio files to be read. If there are, digital mixing program 90 loops to BOX 110 and repeats the operations described above until peak absolute values and RMS values are obtained for all files. When all files have been read and pre-

processed as needed and their characteristic parameters (such as peak absolute value and RMS value) have been determined digital mixing program 90 advances to BOX 180 and calculates scale factors to apply to each file respectively. Digital mixing program 90 then continues to BOX 200.

At BOX 210, digital mixing program 90 reads samples from all input audio files for a second time.

At BOX 220, digital mixing program pre-processes the digital audio files a second time. The pre-processing of BOX 220 may comprise adding reverb, audio compression, applying stereo imaging, applying equalization, and pitch correction to the audio file. As before, it may be that not all audio files will require pre-processing; any files intended for pre-processing in the earlier stages of the program are pre-processed now.

At BOX 230 sample rates of the audio files are converted as needed to bring all audio data to a common sample rate using one of many methods commonly known in the art. The target sample rate is typically the highest rate

among the input audio files, though it may be desirable in some instances to choose a lower target sample rate.

At BOX 240 each resulting audio file sample is multiplied by its respective scale factor and then at BOX 250 time-corresponding samples are summed to create a single sample set. At BOX 260 the single sample set is written to a single output file and digital audio program 90 stops.

Referring now to FIG. 2A, there is shown an expanded functional block diagram illustrating digital mixing program 90, more specifically illustrating the first functional block 100 of digital mixing program 90.

At BOX 105, digital mixing program 90 determines the number of audio files (*N*).

At BOX 107 a file pointer variable *i* is set equal to *1*.

At BOX 110 the digital mixing program 90 reads the header of file *i* (initially set to *1*) to determine its type, including whether it is in a compressed format, its sample rate, duration, imaging (stereo or monaural), and any other relevant data contained in the file header.

At Diamond 120 it is determined whether audio file *i* is in a compressed format. If the digital audio file is in a compressed format then at BOX 130 the file is expanded into an uncompressed format and the process advances to Node 133. If the digital audio file is in an uncompressed format then digital mixing program 90 advances to Node 133.

At BOX 135 digital mixing program 90 initializes variables *PEAKREG* and *SUMREG* by setting each variable equal to zero.

At BOX 140, digital mixing program 90 reads the first sample and in subsequent loops reads the next consecutive sample contained within audio file *i*.

At BOX 150 the current sample of file *i* undergoes pre-processing. Pre-processing may comprise adding reverb to the audio file, applying audio compression, applying stereo imaging, applying equalization, and applying pitch correction to the audio file. It may be that not all files require pre-processing.

At BOX 152 digital mixing program 90 determines if the absolute value of the current pre-processed sample is greater than the value last assigned to *PEAKREG*. If the absolute value of the current pre-processed sample is greater than the current value of *PEAKREG*, then *PEAKREG* is set equal to the absolute value of the current pre-processed sample.

At BOX 154, digital mixing program 90 sets the value of SUMREG equal to the current value of SUMREG plus the square of the current pre-processed sample value.

At Diamond 156 it is determined whether any samples remain within audio file *i*. If samples remain then digital mixing program 90 loops back to BOX 140

and the process described above is repeated. If no samples remain within the digital audio file then the process advances to BOX 160.

At BOX 160 the peak absolute value of file *i* (*PEAKi*) is determined to be the current value of *PEAKREG* and the root mean square (RMS) value for file *i* (*RMSi*) is calculated from the current value of *SUMREG* according to the formula below.

$$RMSi = SQRT(SUMREG/N_{samples})$$

At BOX 168 digital mixing program increments the value of *i*. The value of *i* is incremented according to the following equation.

$$i = (i+1)$$

At Diamond 170 it is determined whether *i* is greater than *N*. If *i* is not greater than *N* then the process advances to Node 109 and the process described above is repeated starting with BOX 110 and the next audio file is processed. If *i*

is greater than N then all files have been processed and the process advances to BOX 180.

At BOX 180 the scale factors for each audio file i are calculated. For example, suppose there are two audio files, the first file being monaural and the second being stereo, at BOX 180 two separate scale factors would be calculated. A first scale factor for the first audio file is calculated for later application to samples of the first audio file. A second scale factor is calculated for the second audio file for later application to samples of the right and left channels of the second audio file. This can be more easily understood with reference to FIGS. 3A and 3B. Referring now to FIG. 3A equations are shown for determining the scale factors for two audio files given their peak absolute values, *PEAK1* and *PEAK2*, and their RMS values, *RMS1* and *RMS2*, a mixing factor, β, and a constant value, K. The mixing factor, β, may take on values from zero to one but is typically set to *0.5*. The constant, K, is the maximum sample value allowed by the output audio file format. Referring now to FIG. 3B a matrix equation is shown for relating the scale factors, S_i, for N number of audio files to the peak absolute values, P_i, and RMS values, R_i, of the files, mixing factors, β_i, and the constant, K. The mixing

factors, β_i, may take on values from zero to one, so long as their sum is equal to one; K is defined as above. The scale factors are calculated by inverting the matrix equation by any of several methods commonly known in the art.

The process described in FIG. 2A and subsequent figures below may be accomplished by various other similar means. For example, the pre-processed samples created in BOX 150 of FIG. 2A were used only for calculating the peak absolute values and RMS values of the pre-processed audio data sets and were then discarded. The completion of the mixing process requires the pre-processing step to be repeated, as will be shown below.

Referring now to FIG. 2B, there is shown an alternative embodiment of the process of FIG. 2A. The alternative embodiment depicted in FIG. 2B utilizes many of the processes described above with regard to FIG. 2A; therefore, the numbers depicting the process steps in FIG. 2B correspond to those in FIG. 2A and the description given above. With regard to the process of FIG. 2B, the processes having the same number as those described in reference to FIG. 2A are identical, except that an additional process has been added at BOX 155 in which

pre-processed samples are saved to a temporary file for later use. An individual temporary file is required for saving each pre-processed file. For most pre-processing algorithms implemented in most computing systems, the time required to repeat pre-processing is far less than the time required to write and read back a temporary file, so the embodiment of FIG. 2A is preferred over that of FIG. 2B.

Referring now to FIG. 2C, there is shown a second alternative embodiment for the process of FIG. 2A is shown in FIG. 2C. With regard to the process of FIG. 2C, the same reference numbers of FIG. 2A have been utilized to denote processes that are identical in function and description. In this embodiment pre-processing is not performed on any file during the file reading and analysis stage. The peak absolute values and RMS values are calculated from all audio files in their unprocessed states. Instead of pre-processing first, the effects of later pre-processing are estimated and the calculated peak absolute values and RMS values are modified based on the predetermined estimate. The effects of preprocessing are predetermined by doing statistical and psychoacoustic testing to assess the effects of preprocessing on the peak absolute value, RMS value or other file characteristics of typical audio files. After file characteristics are determined they

are modified to emulate the effects of pre-processing. For example, suppose that reverberation pre-processing is to be applied to a particular file before the final scaling and summation step, and it is known that the reverberation pre-processing generally increases an audio file's peak absolute value and RMS value by 50%. Then the peak absolute value and RMS value for the file to be pre-processed, calculated in BOX 160, are modified in BOX 175 by multiplying them by a factor of *1.5*. The method of FIG. 2C is the most efficient of the three methods described, but introduces uncertainty in determining the scale factors unless the subsequent pre-processing algorithm is very well characterized.

Digital mixing program 90 then advances to Node 181. From Node 181, digital mixing program 90 advances to the digital audio summation program 200, illustrated previously in a simplified view in FIG. 1. The process could be accomplished as described in FIG. 1, but it would be very inefficient. Accordingly, the preferred embodiment of the invention utilizes the more efficient method described in relation to FIG. 4.

Referring now to FIG. 4, there is shown a preferred embodiment of BOX 200 of FIG. 1. This embodiment handles the case where only two files are to be summed and one file's sample rate is exactly twice that of the other. This is not intended to be limiting in any way and it will be clear to those skilled in the art that these techniques may be expanded to sum a larger group of files with various sampling rates.

At BOX 300 the first samples of each audio file are read, and the files are temporally aligned. At BOX 310 the pre-processing is applied to the samples if required.

At Diamond 320 it is determined whether there are two aligned samples to sum together. If there are two samples, digital mixing program 90 advances to BOX 330 where each of the samples is multiplied by its respective scale factor, calculated during the process of BOX 100, then at BOX 340 the samples are summed. This process is performed for monaural and stereo files, though for stereo files, corresponding left channel samples are scaled and summed and corresponding right channel samples are scaled and summed to create left and

right output samples, respectively. Typically, for the combination of a stereo and a mono file, samples from the mono file are scaled and summed equally with corresponding scaled right and left samples of the stereo file to create right and left output samples, respectively. Digital mixing program 90 advances to BOX 350 where the summed samples from BOX 340 are saved in a single digital audio file.

At Diamond 360 the input files are examined to determine if any samples remain. If so, digital mixing program 90 advances to BOX 370 where the next samples are read. Then the digital mixing program 90 returns execution to BOX 310.

If at Diamond 320 there were not two aligned samples, the digital mixing program 90 would advance to BOX 380 to generate data for the missing sample utilizing the following process. At BOX 380, digital mixing program 90 acquires the samples preceding and succeeding the missing sample, and at BOX 390 the preceding and succeeding samples are summed and then multiplied by a factor of 1/2 to generate an interpolated sample. This process is undertaken for both the right and left channels if the audio file is stereo. The interpolated sample aligns

with the sample from the other audio file and the samples are scaled when execution continues at BOX 330.

At Diamond 360, if it is found that one audio file has greater length than the other audio file, the shorter audio file is lengthened to match the other file by appending zero-valued samples to the shorter file. If no more samples remain in either file, the mixing process is complete and execution stops.

If the process of BOX 100 in FIG. 2 was accomplished according to the method of FIG. 2B, then in BOX 300 and BOX 370 of FIG. 4 samples are read from the temporary audio data files created in BOX 155 of FIG. 2B and the pre-processing step in BOX 310 of FIG. 4 is omitted.

Although the present invention has been described as being applied to two audio files with a two-to-one sample rate ratio, the present invention may be applied to *N* number of audio files with any combination of sample rates, the rates converted to a single common sample rate by any one of several commonly known methods. Additionally, the audio files utilized by the present invention may be

either stereophonic or monaural. The present invention may be embodied in a client server device operatively coupled over a network for communication.

Also, although the present invention has been described with reference to an implementation utilizing the main processor of a personal computer, it will be clear to those skilled in the art that it could be implemented as a dedicated hardware subsystem with the functions described above instantiated in firmware. The resulting hardware subsystem could take the form of a dedicated digital signal processing module embedded in a server computer or a client computer or a stand-alone recording and playback device.